Stat 564
10-12-17

ANOVA table for multiple regression

(1)

Source	SS	df	MS	F
REG	SSR	k	SSR/k	MSR/MSE
ERR	SSE	$n-p$	$\frac{SSE}{n-p}$	
TOT	SST	$n-1$		

$SSE = Y'(I-H)Y$

$p = k+1$

$F_{k,n-p}$

The F test is for

$H_0: \beta_1 = \beta_2 = \dots = \beta_k = 0$

H_1: At least one $\beta_i \neq 0$

(2)

$$SST = S_{yy} = \sum_{i=1}^{n}(y_i - \bar{y})^2$$

$$= [y_1-\bar{y}, y_2-\bar{y}, \dots, y_n-\bar{y}]\underbrace{\begin{bmatrix} y_1-\bar{y} \\ y_2-\bar{y} \\ \vdots \\ y_n-\bar{y} \end{bmatrix}}_{Y-\bar{y}\vec{1}}$$

$$SST = (Y-\bar{y}\vec{1})'(Y-\bar{y}\vec{1})$$

$MSE = \hat{\sigma}^2$ is still an unbiased estimator of σ^2

$$\hat{\sigma} = \sqrt{MSE} = s$$

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST}$$

= portion of the total variation in y that is explained by the predictors

$$R^2_{adj} = 1 - \frac{MSE}{MST} = 1 - \frac{SSE/(n-p)}{SST/(n-1)}$$

$$= 1 - \frac{SSE}{SST} \cdot \frac{n-1}{n-p}$$

4

Hypothesis tests & confidence intervals for any linear combination of parameters

Let $\theta = a_0\beta_0 + a_1\beta_1 + \dots + a_k\beta_k$

$$= [a_0, a_1, \dots, a_k]\begin{bmatrix}\beta_0\\ \beta_1\\ \vdots\\ \beta_k\end{bmatrix} = \vec{a}'\beta$$

Let $\hat{\theta} = a'\hat{\beta}$. Then $E(\hat{\theta}) = a' E(\hat{\beta}) = a'\beta = \theta$

(5)

$$\text{Cov}(\hat{\theta}) = \text{Cov}(a'\hat{\beta}) = a' \underbrace{\text{Cov}(\hat{\beta})}_{\sigma^2 (X'X)^{-1}} a$$

$$\text{Cov}(\hat{\theta}) = \sigma^2 a'(X'X)^{-1} a$$

$$\text{Standard error of } \hat{\theta} = \sqrt{MSE \cdot a'(X'X)^{-1}a}$$

$$= s\sqrt{a'(X'X)^{-1}a}$$

C.I. for θ:

$$\hat{\theta} \pm t_{\alpha/2}\, s\sqrt{a'(X'X)^{-1}a}$$

$\uparrow$ df $= n-p$

(6)

To test $H_0: \theta = \theta_0$

$H_1: \theta \neq \theta_0$

Test stat is

$$\frac{\hat{\theta} - \theta_0}{s\sqrt{a'(X'X)^{-1}a}}$$

This is a t-test with $n-p$ df

Special case: $\theta = \beta_i$

$$a = \begin{bmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix} \leftarrow i^{th}$$

⑦

Then $a'(X'X)^{-1}a$ will simply pick out the ii^{th} (diagonal) element of $(X'X)^{-1}$

$H_0: \beta_i = c$

$H_1: \beta_i \neq c$

$$\text{Test stat} = \frac{\hat{\beta}_i - c}{s\sqrt{ii^{th} \text{ element of } (X'X)^{-1}}}$$

Another special case. $H_0: \beta_1 = \beta_2 \quad (\beta_1 - \beta_2 = 0)$

$H_1: \beta_1 \neq \beta_2 \quad (\beta_1 - \beta_2 \neq 0)$

⑧

$\theta = \beta_1 - \beta_2$ so $a = \begin{bmatrix} 0 \\ 1 \\ -1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$

$$\text{Test stat} = \frac{\hat{\beta}_1 - \hat{\beta}_2 - 0}{s\sqrt{a'(X'X)^{-1}a}}$$

Let $\vec{\theta} = A\beta$ then $\hat{\theta} = A\hat{\beta}$

$E(\hat{\theta}) = A\,E(\hat{\beta}) = A\beta = \theta$

$Cov(\hat{\theta}) = Cov(A\hat{\beta}) = A\,Cov(\hat{\beta})A'$

9

Special case: $H_0: \beta_1 = \beta_2 = \beta_3$

$$\begin{bmatrix} 0 & 1 & -1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & -1 & 0 & \cdots & 0 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \vdots \\ \beta_k \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$A \qquad \beta \quad = \quad C$$

10

Full & reduced model
or
restricted

Full model $y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k + \varepsilon$

$H_0: \beta_1 = \beta_2 = \beta_3$

Restricted model $y = \beta_0 + \beta_1 X_1 + \beta_1 X_2 + \beta_1 X_3 + \beta_4 X_4 + \cdots + \varepsilon$

The additional sum of squares F test:

(11)

$$F = \frac{\left[\dfrac{SSE_{red} - SSE_{full}}{dfE_{red} - dfE_{full}}\right]}{MSE_{full}}$$

$SSE_{red} - SSE_{full}$ ← additional sum of squares

$dfE_{red} - dfE_{full}$ ← rank of A

Method: ① Run the full model & get SSE_{full}, dfE_{full}, MSE_{full}

② Run the reduced model & get SSE_{red}, dfE_{red}

Assemble F.

(in SAS, see PROC REG & TEST statement)

(12)

The Bonferroni principle

Let $E_1, \ldots, E_n$ be n events whose probabilities are $p_1, \ldots, p_n$

$$P(E_1 \cap \ldots \cap E_n) = 1 - P(\overline{E_1 \cap \ldots \cap E_n})$$

$$= 1 - P(\overline{E}_1 \cup \ldots \cup \overline{E}_n) \quad \text{deMorgan's Law}$$

$$\geq 1 - P(\overline{E}_1) + \ldots + P(\overline{E}_n)$$

$$= 1 - [(1-p_1) + \ldots + (1-p_n)]$$

Application to confidence intervals:

Suppose you do 10 confidence intervals, each 95%

$$P(E_1 \cap \ldots \cap E_{10}) \geq 1 - [.05 + \ldots + .05]$$
$$= .5$$

HW #3 follows this page

3.1 Consider the National Football League data in Table B.1.

a. Fit a multiple linear regression model relating the number of games won to team's passing yardage (x_2), the percentage of rushing plays (x_7), and the op yards rushing (x_8).

b. Construct the analysis-of-variance table and test for significance of regres

c. Calculate t statistics for testing the hypotheses H_0: $\beta_2 = 0$, H_0: $\beta_7 = 0$, and 0. What conclusions can you draw about the roles the variables x_2, x_7, and x_8 the model?

d. Calculate R^2 and R^2_{Adj} for this model.

e. Using the partial F test, determine the contribution of x_7 to the model. Hov partial F statistic related to the t test for β_7 calculated in part c above?

3.3 Refer to Problem 3.1.

a. Find a 95% CI on β_7.

b. Find a 95% CI on the mean number of games won by a team when x_2 = 23 56.0, and $x_8 = 2100$.